DOCUMENT RESUME

ED 128 380 TM 005 502

AUTHOR Halperin, Silas
TITLE Ambiguity in ANOVA: Let the Hypothesis Be Your Guide.
PUB DATE [Apr 76]
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976); Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Analysis of Variance; *Comparative Analysis; Computer Programs; *Hypothesis Testing; *Mathematical Models

ABSTRACT

Although analysis of variance is the most popular statistical tool to researchers in the behavioral sciences, the casual user has recognized only recently that there is no single correct way to perform such an analysis. This paper is concerned with hypotheses tested in the two-way, fixed model under a variety of conditions: (1) with and without restrictions, and (2) with unbalanced, proportional and balanced cell frequencies. Standard texts and canned computer programs are surveyed and their recommendations are expressed in a common notation. Recommendations based on these results are given, along with the rationale for the recommendations. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED128380

BEST COPY AVAILABLE

TM005 502

Ambiguity in ANOVA

Let the Hypothesis be Your Guide

Silas Halperin

Syracuse University

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1976.

INTRODUCTION

Although analysis of variance is the most popular statistical tool for researchers in the behavioral sciences, it is only recently that the casual user has recognized that there is no single correct way to perform such an analysis. For a given factorial design with equal cell frequencies, older textbooks uniformly described exactly the same procedure. Rarely was the topic of unbalanced data discussed.

Within the past decade, there has developed an increased awareness that analysis of variance can be viewed as a special case of regression analysis. This more general approach has made it clear that, for unbalanced designs, there is no unique way to perform analysis of variance. For example, Overall and Klett (1972) and Kerlinger and Pedhazur (1973) point out several different ways to calculate a main effect sum of squares. However, they do not provide adequate advice on how to choose among the various sums of squares.

A substantial literature is available to help in the selection of the most appropriate method of analysis. Unfortunately, much of this literature is very esoteric and is thus ineffective as an aid to the non-statistician. It is the purpose of this paper to integrate the literature into a comprehensive comparison of alternative methods of performing a two-way, fixed effects analysis of variance.

THE MODEL

The two-way (AxB) analysis of variance is based on the model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \quad , \qquad (1)$$

where $i=1,\ldots, a$; $j=1,\ldots, b$; and $k=1,\ldots,n_{ij}$. Here, the $n_{ij}>0$ denotes the cell frequencies.

Using the notation of Searle (1971), we can denote as $R(\mu, \alpha,\beta, \gamma)$ the reduction in the sum of squares due to the model in (1). Similarly, we could define a reduced model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \qquad (2)$$

and $R(\mu, \alpha, \beta)$ would then denote the reduction in the sum of squares due to this new model. The difference between these reductions is denoted

$$R(\gamma \mid \mu, \alpha, \beta) = R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha, \beta) \qquad (3)$$

and expresses the reduction due to fitting γ over and above μ, α, and β. This would commonly be considered as the sum of squares due to interaction.

In an analogous fashion, we could calculate $R(\alpha \mid \mu, \beta, \gamma)$, $R(\alpha \mid \mu, \beta)$ or $R(\alpha \mid \mu)$. Any, or all of these could, under the correct conditions, be interpreted as a sum of squares for the "A" main effect. How is one to choose? What are the "correct conditions" under which all of these may be interpreted as a SS(A)?

The answers to these questions lie deeply imbedded in the interpretations one makes of the parameters in the full model (1). It is a common misconception that μ, α_i, β_j and γ_{ij} must be interpreted as a grand mean, A effect, B effect and interaction, respectively. Under certain restrictions on the model, these interpretations are correct. However, such restrictions need not be imposed to derive the analysis.

THE UNRESTRICTED MODEL

The analysis of variance based on the model (1) with no restrictions

imposed is fully developed by Searle (1971). Since the normal equations for the unrestricted model have no unique least squares solution, Searle makes use of the concept of a generalized inverse.

Using matrix notation, the model can be re-expressed as

$$\underline{y} = X\underline{\theta} + \underline{e}, \qquad (4)$$

where $\underline{y}$ is an Nx1 vector of the y_{ijk}, $\underline{\theta}$ contains the parameters $\left\{\mu, \alpha_i, \beta_j, \gamma_{ij}\right\}$, $\underline{e}$ is an Nx1 vector of the e_{ijk} and X is a design matrix consisting entirely of zeros and ones. Normal equations can now be expressed as

$$(X'X)\,\hat{\underline{\theta}} = X'\,\underline{y}\,, \qquad (5)$$

where $\hat{\underline{\theta}}$ represents some least squares solution. Since $X'X$ is singular, we resort to a generalized inverse, G, which satisfies

$$(X'X)\,G\,(X'X) = X'X\,, \qquad (6)$$

so that

$$\hat{\underline{\theta}} = G\,X'\,\underline{y}\,. \qquad (7)$$

Searle makes elegant use of the properties of generalized inverses to show that the reduction due to the model is invariant to the choice of G. However, choosing a specific G is equivalent to placing constraints on the solution $\hat{\underline{\theta}}$. It is critical at this point to recognize that the choice of constraints on the solution is a matter of convenience and in no way forces similar restrictions on the parameters of the model.

Using generalized inverses to solve normal equations for the full and a variety of reduced models, it is simple to derive such expressions as $R_o(\gamma \mid \mu, \alpha, \beta)$, $R_o(\alpha \mid \mu, \beta, \gamma)$, $R_o(\alpha \mid \mu, \beta)$ and $R_o(\alpha \mid \mu)$. The subscript on $R_o(\ )$ will indicate a reduction based on an unstricted model. Compu-

tational formulas and hypotheses tested by these sums of squares are given in Table 1.

Insert Table 1 about here

Observe one strange occurrence in Table 1:

$$R_o(\alpha|\mu, \beta, \gamma) = 0$$

This startling outcome results from the fact that the design matrices for the full model and for the reduced model

$$Y_{ijk} = \mu + \beta_j + \gamma_{ij} + e_{ijk} \qquad (8)$$

span exactly the same vector space. In other words, since the full model has far more parameters than can be used, the elimination of the parameters α_i does not reduce the effectiveness of the model. In fact, the elimination of μ, α_i and β_j would not change the reduction in the sum of squares due to the model. Thus, the reduction due to any effect, over and above the γ_{ij}, is uninteresting in the unrestricted model.

Again referring to Table 1, note that the hypotheses being tested are complex and, most likely, not very interpretable. Furthermore, the hypotheses corresponding to $R_o(\alpha|\mu)$ and $R_o(\alpha|\mu,\beta)$ depend upon the possibly arbitrary configuration of cell frequencies. These findings provide little direction in the choice of a sum of squares for the A-effect. Is it possible that some form of restrictions imposed on the model might simplify these hypotheses? Will imposition of these restriction change the reduction sums of squares? What type of restrictions should be chosen?

THE RESTRICTED MODEL

To answer these questions, the model under restrictions will be investigated. Restrictions can be imposed in a very general way: for an arbi-

trary set of numbers $\{v_i;\ i=1,\ldots,a\}$ and $\{w_j;\ j=1,\ldots,b\}$ where $\Sigma v_i = \Sigma w_j = 1$, we can impose restrictions of the form

$$\sum_i v_i\alpha_i = \sum_j w_j\beta_j = \Sigma v_i\gamma_{ij} = \Sigma w_j\gamma_{ij} = 0 \qquad (9)$$

Denoting as $R_r(\ )$ the reduction in the sum of squares due to a model restricted under (9), it is possible to show, using results from Scheffé (1959) that $R_r(\gamma|u,\alpha,\beta) = R_o(\gamma|u,\alpha,\beta)$ and $R_r(\alpha|u,\beta) = R_o(\alpha|u,\beta)$. Although Scheffé does not consider the situation, it is easily shown that $R_r(\alpha|u) = R_o(\alpha|u)$. However, Scheffé does show that, in general, $R_r(\alpha|u,\beta,\gamma)$ does not equal zero as does $R_o(\alpha|u,\beta,\gamma)$; furthermore, $R_r(\alpha|u,\beta,\gamma)$ depends upon the choice of the weights $\{v_i, w_j\}$. For this reason, two commonly used sets of restrictions will be explored. The first set of restrictions, defined by the weights given in set S_1, are

$$S_1 = \left\{v_i = \frac{1}{a},\ w_j = \frac{1}{b}\right\}. \qquad (10)$$

This implies that

$$\alpha_\cdot = \beta_\cdot = \gamma_{i\cdot} = \gamma_{\cdot j} = 0 \qquad (11)$$

where the dot notation signifies summing over the subscript replaced by the dot. A second restriction is defined by the weights

$$S_2 = \left\{v_i = n_{i\cdot}/N\ ,\ w_j = n_{\cdot j}/N\right\}, \qquad (12)$$

which in turn imposes the restrictions

$$\Sigma n_{i\cdot}\ \alpha_i = \Sigma n_{\cdot j}\beta_j = \Sigma n_{i\cdot}\ \gamma_{ij} = \Sigma n_{\cdot j}\ \gamma_{ij} = 0. \qquad (13)$$

It should be clear that the manner in which the parameters of the model relate to the cell means, u_{ij}, depends upon the choice of a restriction. When no restrictions are imposed, no unique functional relationship exists

between the model's parameters and the cell means. It is informative to express at least one set of parameters, say the α_i, in terms of the cell means to see the implications of the two restrictions. Under S_1, we find that

$$\alpha_i^{(1)} = \frac{1}{b}\sum_j \mu_{ij} - \frac{1}{ab}\sum_i\sum_j \mu_{ij} = \pi_{i.} - \pi \ ;$$

Under S_2, the α_i are defined

$$\alpha_i^{(2)} = \frac{1}{N}\sum_j n_{.j}\mu_{ij} - \frac{1}{N}\sum_i\sum_j n_{i.}n_{.j}\mu_{ij} \ .$$

Situations obviously exist where the $\alpha_i^{(1)}$ are all zero, but the $\alpha_i^{(2)}$ are not. For this reason, caution must be exercised when testing a hypothesis of the form

$$H_o: \alpha_i = 0 \text{ for all } i \ .$$

Scheffé proves that, for any set of weights,

$$R_r(\alpha \mid \mu,\beta,\gamma) = \sum_i W_i \hat{A}_i^2 - \left(\sum_i W_i\right)^{-1}\left(\sum_i W_i\hat{A}_i\right)^2, \qquad (14)$$

where $W_i = \left[\sum_j w_j^2/n_{ij}\right]^{-1}$ and $\hat{A}_i = \sum_j w_j \bar{y}_{ij.}$. Substituting S_1 in (14) and denoting the reduction as $R_1(\ )$, simple algebra yields

$$R_1(\alpha \mid \mu,\beta,\gamma) = \sum_i \frac{\left(\sum_j y_{ij.}\right)^2}{\sum_j 1/n_{ij}} - \frac{\left(\sum_i\sum_j \bar{y}_{ij.}\right)^2}{\sum_i\sum_j 1/n_{ij}} \ .$$

This sum of squares is identical to that calculated by the weighted squares of means method (see Searle (1971, pp. 369-372) or Winer (1971, pp. 417-418)).

Applying frequency weighted restrictions as given in S_2, we obtain

from (14) that

$$R_2(\alpha \mid \mu,\beta,\gamma) = \sum_i \frac{(\sum_j n_{.j}\bar{y}_{ij.})^2}{\sum_j (n_{.j}^2/n_{ij})} - \frac{\left(\sum_i \sum_j n_{.j}\bar{y}_{ij.}\right)^2}{\sum_i (\sum_j n_{.j}^2/n_{.j})} \quad . \qquad (16)$$

This sum of squares is not in common use. The formulas for $R_r(\gamma|\mu,\alpha,\beta)$ $R_r(\alpha \mid \mu)$, $R_r(\alpha|\mu,\beta)$ and $R_r(\alpha \mid \mu,\beta,\gamma)$ and the hypotheses they test are given, for both S_1 and S_2, in Table 2. Parameters are superscripted to reinforce their differences.

Insert Table 2 about here

Several things should be noted. First, the definitions of the parameters are not the same under S_2 as under S_1. Thus, even when a hypothesis under S_1 appears to be identical to that under S_2, the two may not be equivalent (i.e., imply each other). Scheffé (1959) shows that the interaction hypotheses are equivalent under the two sets. Assuming zero interactions, the hypothesis for $R_1(\alpha|\mu)$ implies that for $R_2(\alpha|\mu)$. Similarly, the two hypotheses for $R_r(\alpha|\mu,\beta)$ are conditionally equivalent. However, the hypotheses for $R_r(\alpha|\mu,\beta\gamma)$ are not equivalent and this is rei. orced by the dependence of $R_r(\alpha|\mu,\beta,\gamma)$ on the weights $\{v_i\}$ and $\{w_j\}$.

It should be noted that $R_r(\alpha|\mu)$ tests a hypothesis which remains dependent upon cell frequencies and the hypothesis for $R_r(\alpha|\mu,\beta)$ simplifies only when expressed conditionally. In summary, imposing restrictions does simplify the hypotheses in Table 1, but the appropriate choice of a main effect SS remains unclear.

CELL MEAN MODEL

If μ_{ij} denotes a cell mean, the model in (1) can be expressed

$$E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} = \mu_{ij} \qquad (17)$$

In the absence of restrictions, μ, α_i, β_j and γ_{ij} have no unique dependence upon the μ_{ij}. Imposing restrictions of the form (9) allows one to express these parameters as functions of the cell means. However, some statisticians prefer to eliminate the problem of imposing restrictions by writing the model entirely in terms of cell means. This form of the model actually predates the overparameterized version.

The cell mean model is discussed briefly by Searle (1971) and in great depth by Timm and Carlson (1973). In a short and well written article by Kutner (1974), the hypotheses and corresponding SS of primary interest are developed. Table 3 presents some of Kutner's null hypotheses, along with their equivalent reductions.

Insert Table 3 about here

Hypothesis H_1 is equivalent to the hypothesis tested by $R_o(\gamma|\mu,\alpha,\beta)$. Hypothesis H_4 is simply a re-writing of that tested by $R_o(\alpha|\mu)$. The reduction $R_o(\alpha|\mu,\beta)$ can be viewed as testing the conditional hypothesis

$$H : \alpha_i \text{ are all equal} \mid \gamma_{ij} \text{ are all equal,}$$

and this is equivalent to H_3 above. The hypothesis tested by $R_1(\alpha|\mu,\beta,\gamma)$ is easily shown to be equivalent to H_2. Searle (1971, p.315, eq. 122) also derives a statistic, equal in value to $R_1(\alpha|\mu,\beta,\gamma)$, but not readily expressed as a reduction. In the unrestricted model, the statistic tests the hypothesis

$H: \alpha_i + \overline{\gamma}_{i.}$ are all equal.

PROPORTIONAL AND BALANCED DATA

It should be clear that $R(\gamma|\mu,\alpha,\beta)$ is our only rational choice for the interaction SS. Of the possible choices of the SS(A), $R(\alpha|\mu)$ appears to be clearly undesirable. The other two reductions, $R(\alpha|\mu,\beta)$ and $R(\alpha|\mu,\beta,\gamma)$ each have drawbacks; $R(\alpha|\mu,\beta)$ tests a very complicated hypothesis in the overparameterized, un-restricted model. In the restricted or in the cell mean model, it tests a conditional hypothesis. The reduction $R(\alpha|\mu,\beta,\gamma)$ tests nothing in the unrestricted model and tests different hypotheses and assumes different values in the restricted models, depending upon the restriction. How do these different reductions operate for data which are proportional?

Cell frequencies are said to be proportional when

$$n_{ij} = \frac{n_{i.}\, n_{.j}}{N} \tag{18}$$

In this situation, we find that

$$\begin{aligned} R_o(\alpha|\mu) = R_1(\alpha|\mu) = R_2(\alpha|\mu) = \\ R_o(\alpha|\mu,\beta) = R_1(\alpha|\mu,\beta) = R_2(\alpha|\mu,\beta) = \\ R_2(\alpha|\mu,\beta,\gamma) \neq R_1(\alpha|\mu,\beta,\gamma)\ . \end{aligned} \tag{19}$$

These reductions and their hypotheses are given in Table 4.

Insert Table 4 about here

Notice that $R(\alpha|\mu)$ and $R(\alpha|\mu,\beta)$ are equal, regardless of which restrictions are imposed or even when no restrictions are imposed. The reduction $R_2(\alpha|\mu,\beta,\gamma)$ equals this common value, but $R_1(\alpha|\mu,\beta,\gamma)$ does <u>not</u>.

The proportionality of the cell frequencies does not substantially simplify the hypotheses tested by $R_o(\alpha|\mu)$ and $R_o(\alpha|\mu,\beta)$. The same is true for $R_1(\alpha|\mu)$ and $R_1(\alpha|\mu,\beta)$. However, under restrictions S_2, the reductions $R_2(\alpha|\mu)$, $R_2(\alpha|\mu,\beta)$ and $R_2(\alpha|\mu,\beta,\gamma)$ all test the simple hypothesis that the α_i are all zero. The reduction $R_1(\alpha|\mu,\beta,\gamma)$ also tests that the α_i are all zero but the α_i have a different meaning, and thus the hypothesis must be viewed as distinct from that for $R_2(\alpha|\mu,\beta,\gamma)$.

The common value of the reductions given in (19) will equal the SS(A) calculated from the special formulas for proportional data. Such formulas are given in standard experimental design books such as Kirk (1968, p. 201).

If all cell frequencies are equal ($n_{ij} = n$), we have a balanced design. In that event, we find that

$$R_o(\alpha|\mu) = R_1(\alpha|\mu) = R_2(\alpha|\mu) =$$
$$R_o(\alpha|\mu,\beta) = R_1(\alpha|\mu,\beta) = R_2(\alpha|\mu,\beta) =$$
$$R_1(\alpha|\mu,\beta,\gamma) = R_2(\alpha|\mu,\beta,\gamma).$$

The reduction $R_o(\alpha|\mu,\beta,\gamma)$ remains zero, but all others are equal and identical to the SS(A) given in any elementary statistics book. The restrictions degenerate to the same set and all $R_1(\ )$ and $R_2(\ )$ test the same hypothesis, $\alpha_i = 0$. The $R_o(\alpha|\mu)$ and $R_o(\alpha|\mu,\beta)$ test the hypothesis that the $\alpha_i + \bar{\gamma}_{i\cdot}$ are equal, where $\bar{\gamma}_{i\cdot} = \frac{1}{b}\sum_j \gamma_{ij}$. This information is summarized in Table 5.

Insert Table 5 about here

THE REDUCTION CHOSEN

Elementary statistics texts such as Hays (1973), McNemar (1962), Glass

and Stanley (1970), Edwards (1968) and Guilford and Fruchter (1973) consider only the balanced case with restrictions and thus have no choice to make. Experimental design texts such as Winer(1962) and Kirk (1968) consider unblanced and proportional designs and recommend, in our notation, $R(\gamma|\mu,\alpha,\beta)$ and $R(\alpha|\mu,\beta)$. In his second edition, Winer (1971) also suggests the weighted squares of means solution, $R_1(\alpha|\mu,\beta,\gamma)$. Texts stressing a linear regression approach (e.g., Kelinger and Pedhazur (1973) and Overall and Klett (1972)) suggest the choice be made from among $R_1(\alpha|\mu)$, $R_1(\alpha|\mu,\beta)$ and $R_1(\alpha|\mu,\beta,\gamma)$.

Canned computer programs approach the unbalanced design in a wide variety of ways. Francis (1973) has surveyed a variety of such packages and reports that some, such as CAROLINA (Psychometric Laboratory, University of North Carolina) and OSIRIS (University of Michigan, 1970) calculate $R_1(\alpha|\mu)$ and then $R_1(\beta|\mu,\alpha)$, or vice-versa, depending on the order of input of the factors. A program from North Carolina State (Barr and Goodnight, 1971) calculates $R_1(\alpha|\mu)$ and $R_1(\beta|\mu)$; the interaction is calculated by subtraction, yielding a negative sum of squares for the example cited.

Two programs from the "Biomedical Computer Programs" (Dixon, 1970) provide the user with among the most accurate and general and least expensive means of performing the analysis. BMDX64 (BMD10V) will automatically calculate $R_1(\alpha|\mu,\beta,\gamma)$, $R_1(\beta|\mu,\alpha,\gamma)$ and $R_1(\gamma|\mu,\alpha,\beta)$. If the user wishes, any other reduction can be obtained by means of extra hypotheses cards. BMD05V is somewhat less automatic and somewhat less general. It requires that the user input a design matrix and the specific reductions required. In light of the wide diversity of possible ways to analyze unbalanced data, it is this author's belief that the less automated approach of BMD05V is to

be preferred, in that it forces the user to be aware of what procedure is being used.

A word of caution is in order at this point. A variety of computer programs are in common use which claim to handle unbalanced data. Many of these provide little (or no) documentation as to the method of analysis. Some, such as AVAR23(Veldman, 1967) in fact perform an unweighted means analysis of variance. This is not an exact least squares solution except in the balanced case. It is always desirable, prior to relying on any computer program, to submit a test problem to it. Overall and Klett (1972) provide one such set of test data (see the footnote in Kutner (1974) for a transcribing error) along with a large variety of different reductions.

RECOMMENDATIONS

Only one definitive recommendation will be offered: use $R(\gamma|\mu,\alpha,\beta)$ as the interaction SS. This is no major breakthrough, since most use this presently. In attempting to choose between reductions for the "A" main effect, this author generally prefers $R(\alpha|\mu,\beta)$. This preference is not based upon desirable properties of $R(\alpha|\mu,\beta)$, but rather on undesirable properties of its competitors. The reduction $R(\alpha|\mu)$ tests hypotheses which are dependent upon cell frequencies for all but the trivial case of equal frequencies and the case of proportional frequencies with weighted restrictions. The reduction $R(\alpha|\mu,\beta,\gamma)$ is widely used, but has the disadvantage of being dependent upon the form of the restrictions placed on the model. One could argue that, if we always choose unweighted restrictions of the form of (11), this problem would not exist. However, we then are forced to use an "unusual" (by widely used textbooks) statistic for the main effects when the cell sizes are proportional. The preferred reduction, $R(\alpha|\mu,\beta)$, always

tests hypotheses which are free of cell frequencies (at least when there are no empty cells). Its value is not affected by restrictions and its hypothesis is always reasonable, even with no restrictions. It is true that it tests a hypothesis which is conditional upon "additivity." This, however, is no real disadvantage, since interest in a "main effect" is typically present only in the absence of interaction. Given even the slightest hint of unequal γ_{ij}, most researchers will go immediately to simple effect

In summary, $R(\alpha|\mu,\beta)$ is the preferred choice of this author. That choice was made on the basis of subjective evaluation. It is incumbent upon each individual to make his or her own decisions.

REFERENCES

Barr, A. J. and Goodnight, J. H. Statistical analysis system. North Carolina State University, Raleigh: Student Supply Store, 1971.

CAROLINA: Multivariate analysis of variance program. University of North Carolina: Psychometric Laboratory.

Dixon, W. J. (Ed.), BMD: Biomedical computer programs. Berkeley: University of California Press, 1970.

Edwards, A. L. Experimental design in psychological research. (3rd. ed.) New York: Holt, Rinehart and Winston, 1968.

Francis, I. A comparison of several analysis of variance programs. Journal of the American Statistical Association, 1973, 68, 860-865.

Glass, G. V. and Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Guilford, J. P. and Fruchter, B. Fundamental statistics in psychology and education. (5th ed.) New York: McGraw-Hill, 1973.

Hays, W. L. Statistics for the social sciences. (2nd. ed.) New York: Holt, Rinehart and Winston, 1973.

Kerlinger, F. N. and Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart and Winston, 1973.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole, 1968.

Kutner, M. H. Hypothesis testing in linear models (Eisenhart Model I). The American Statistician, 1974, 28, (3) 98-100.

Mc Nemar, Q. Psychological statistics. (3rd. ed.) New York: Wiley, 1962.

OSIRIS II, OS Users manual. University of Michigan, Ann Arbor: Inter-University Consortium for Political Research, The Institute for Social Research, 1970.

Overall, J. E. and Klett, C. J. Applied multivariate analysis. New York: Mc Graw-Hill, 1972.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Searle, S. R. Linear models. New York: Wiley, 1971.

Timm, N. H. and Carlson, J. E. Analysis of variance through full rank models. Working Paper No. 21, Department of Education Research, University of Pittsburgh, 1973.

Veldman, D. J. Fortran programming for the behavioral sciences. New York: Holt, Rinehart and Winston, 1967.

Winer, B. J. Statistical principles in experimental design. New York: Mc Graw-Hill, 1962.

Winer, B. J. Statistical principles in experimental design. (2nd. ed.) New York: Mc Graw-Hill, 1971.

TABLE 1

Reductions and Hypotheses for the Unrestricted Model

Reduction	Formula	Hypothesis
$R_o(\gamma \mid \mu,\alpha,\beta)$	$\sum_i \sum_j n_{ij}\bar{y}_{ij.}^2 - \sum_i n_{i.}\bar{y}_{i..}^2 - \underline{r}'\, C^{-1}\, \underline{r}$ where $\underline{r} = (r_j) = (y_{.j.} - \sum_i n_{ij}\bar{y}_{i..})$ $C = (C_{jj'}) = (\delta_{jj'} n_{.j} - \sum_i n_{ij} n_{ij'}/n_{i.})$	$\gamma_{ij} = \bar{\gamma}_{..}$ for all i,j where $\bar{\gamma}_{..} = \frac{1}{ab}\sum_i \sum_j \gamma_{ij}$
$R_o(\alpha \mid \mu,\beta,\gamma)$	0	none
$R_o(\alpha \mid \mu,\beta)$	$\sum_i n_{i.}\bar{y}_{i..}^2 + \underline{r}'C^{-1}\underline{r} - \sum_j n_{.j}\bar{y}_{.j.}^2$ where $\underline{r}$ and C are defined as for $R_o(\gamma \mid \mu,\alpha,\beta)$	$\phi_i = 0$; i=1,2,..., a-1 where $\phi_i = (n_{i.} - \sum_j \frac{n_{ij}^2}{n_{.j}})\alpha_i - \sum_{k\neq i} (\sum_j \frac{n_{ij}n_{kj}}{n_{.j}})\alpha_k +$ $\sum_j (n_{ij} - \frac{n_{ij}^2}{n_{.j}})\gamma_{ij} - \sum_{k\neq i}\sum_j (\frac{n_{ij}n_{kj}}{n_{.j}})\gamma_{kj}$ or $\alpha_i = \bar{\alpha}_.$ for all i $\mid$ $\gamma_{ij} = \bar{\gamma}_{..}$ for all i,j
$R_o(\alpha \mid \mu)$	$\sum_i n_{i.}\bar{y}_{i..}^2 - N\bar{y}_{...}^2$	$\alpha_i + \sum_j \frac{n_{ij}}{n_{i.}}(\beta_j + \gamma_{ij})$ equal for all i

TABLE 2

Reductions and Hypotheses for the Restricted Models

Reduction	Formula	Hypothesis*
$R_1(\gamma\|\mu,\alpha,\beta)$	Same as $R_o(\gamma\|\mu,\alpha,\beta)$	all $\gamma_{ij}^{(1)} = 0$
$R_2(\gamma\|\mu,\alpha,\beta)$	Same as $R_o(\gamma\|\mu,\alpha,\beta)$	all $\gamma_{ij}^{(2)} = 0$
$R_1(\alpha\|\mu,\beta,\gamma)$	$\sum_i \frac{(\sum_j \bar{y}_{ij.})^2}{\sum_j n_{ij}^{-1}} - \frac{\left(\sum_i \frac{\sum_j \bar{y}_{ij.}}{\sum_j n_{ij}^{-1}}\right)^2}{\sum_i (\sum_j n_{ij}^{-1})^{-1}}$	all $\alpha_i^{(1)} = 0$
$R_2(\alpha\|\mu,\beta,\gamma)$	$\sum_i \frac{(\sum_j n_{ij}\bar{y}_{ij.})^2}{\sum_j n_{.j}^2 n_{ij}^{-1}} - \frac{\left(\sum_i \frac{\sum_j n_{.j}\bar{y}_{ij.}}{\sum_j n_{.j}^2 n_{ij}^{-1}}\right)^2}{\sum_i (\sum_j n_{.j}^2 n_{ij}^{-1})^{-1}}$	all $\alpha_i^{(2)} = 0$
$R_1(\alpha\|\mu,\beta)$	same as $R_o(\alpha\|\mu,\beta)$	all $\alpha_i^{(1)} = 0 \mid$ all $\gamma_{ij}^{(1)} = 0$
$R_2(\alpha\|\mu,\beta)$	same as $R_o(\alpha\|\mu,\beta)$	all $\alpha_i^{(2)} = 0 \mid$ all $\gamma_{ij}^{(2)} = 0$
$R_1(\alpha\|\mu)$	same as $R_o(\alpha\|\mu)$	all $(\alpha_i^{(1)} + \sum_j \frac{n_{ij}}{n_{i.}} (\beta_j^{(1)} + \gamma_{ij}^{(1)})_{\text{equal}})$
$R_2(\alpha\|\mu)$	same as $R_o(\alpha\|\mu)$	all $(\alpha_i^{(2)} + \sum_j \frac{n_{ij}}{n_{i.}} (\beta_j^{(2)} + \gamma_{ij}^{(2)})_{\text{equal}})$

* $\gamma_{ij}^{(1)} = \mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \bar{\mu}_{..}$; $\gamma_{ij}^{(2)} = \mu_{ij} - \sum_i \frac{n_{i.}}{N} \mu_{ij} - \sum_j \frac{n_{.j}}{N} \mu_{ij} + \sum\sum_{ij} \frac{n_{i.} n_{.j}}{N^2} \mu_{ij}$

$\alpha_i^{(1)} = \bar{\mu}_{i.} - \bar{\mu}_{..}$; $\alpha^{(2)} = \sum_i \frac{n_{i.}}{N} \mu_{ij} - \sum\sum_{ij} \frac{n_{i.} n_{.j}}{N^2} \mu_{ij}$

TABLE 3

Hypotheses and Sums of Squares for the Cell Mean Model

Hypothesis cell means	Hypothesis overparameterized model	Reduction
$H_1: \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$ for all i,i',j,j' $(i \neq i', j \neq j')$	all $\gamma_{ij} = \bar{\gamma}_{..}$	$R_o(\gamma \mid \mu,\alpha,\beta)$
$H_2: \bar{\mu}_{i.} = \bar{\mu}_{..}$ for all i	all $(\alpha_i + \bar{\gamma}_{i.})$ equal	$R_1(\alpha \mid \mu,\beta,\gamma)$
$H_3: H_2 \mid H_1$	all $\alpha_i = \bar{\alpha}_{.} \mid$ all $\gamma_{ij} = \bar{\gamma}_{..}$	$R_o(\alpha \mid \mu,\beta)$
$H_4: \sum_j (n_{ij}/n_{i.})\mu_{ij}$ are all equal	all $(\alpha_i + \sum_j \frac{n_{ij}}{n_{i.}}(\beta_j + \gamma_{ij}))$ equal	$R_o(\alpha \mid \mu)$

TABLE 4

Reductions and Hypotheses for Proportional Cell Frequencies

Reduction	Formula	Hypothesis
$R_o(\gamma\|\mu,\alpha,\beta)$	$\sum_{ij} n_{ij}\bar{y}_{ij.}^2 - \sum_i n_{i.}\bar{y}_{i..}^2 - \sum_j n_{.j}\bar{y}_{.j.}^2 + \bar{y}_{...}^2$	all $\gamma_{ij} = \bar{\gamma}_{..}$
$R_1(\gamma\|\mu,\alpha,\beta)$	same as $R_o(\gamma\|\mu,\alpha,\beta)$	all $\gamma_{ij}^{(1)} = 0$
$R_2(\gamma\|\mu,\alpha,\beta)$	same as $R_o(\gamma\|\mu,\alpha,\beta)$	all $\gamma_{ij}^{(2)} = 0$
$R_o(\alpha\|\mu,\beta,\gamma)$	0	none
$R_1(\alpha\|\mu,\beta,\gamma)$	does not simplify	all $\alpha_i^{(1)} = 0$
$R_2(\alpha\|\mu,\beta,\gamma)$	$\sum_i n_{i.}\bar{y}_{i..}^2 - N\bar{y}_{...}^2$	all $\alpha_i^{(2)} = 0$
$R_o(\alpha\|\mu,\beta)$	same as $R_2(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i = \bar{\alpha}_{.}$ | all $\gamma_{ij} = \bar{\gamma}_{..}$
$R_1(\alpha\|\mu,\beta)$	same as $R_2(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i^{(1)} = 0$ | all $\gamma_{ij}^{(1)} = \bar{\gamma}_{..}$
$R_2(\alpha\|\mu,\beta)$	same as $R_2(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i^{(2)} = 0$ | all $\gamma_{ij}^{(2)} = \bar{\gamma}_{..}$
$R_o(\alpha\|\mu)$	same as $R_2(\alpha\|\mu,\beta,\gamma)$	all $(\alpha_i + \sum_j \frac{n_{ij}}{n_{i.}}(\beta_j + \gamma_{ij}))$ equal
$R_1(\alpha\|\mu)$	same as $R_2(\alpha\|\mu,\beta,\gamma)$	all $(\alpha_i^{(1)} + \sum_j \frac{n_{ij}}{n_{i.}}(\beta_j^{(1)} + \gamma_{ij}^{(1)}))$ equal
$R_2(\alpha\|\mu)$	same as $R_2(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i^{(2)} = 0$

TABLE 5

Reductions and Hypotheses for Equal Cell Frequencies

Reduction	Formula	Hypothesis
$R_o(\gamma\|\mu,\alpha,\beta)$	$n\sum_{ij}\bar{y}_{ij.}^2 - bn\sum_i \bar{y}_{i..}^2 - an\sum_j \bar{y}_{.j.}^2 + N\bar{y}_{...}^2$	all $\gamma_{ij} = \bar{\gamma}_{..}$
$R_1(\gamma\|\mu,\alpha,\beta)$	same as $R_o(\gamma\|\mu,\alpha,\beta)$	all $\gamma_{ij} = 0$
$R_2(\gamma\|\mu,\alpha,\beta)$	same as $R_o(\gamma\|\mu,\alpha,\beta)$	all $\gamma_{ij} = 0$
$R_o(\alpha\|\mu,\beta,\gamma)$	0	none
$R_1(\alpha\|\mu,\beta,\gamma)$	$bn\sum_i \bar{y}_{i..}^2 - N\bar{y}_{...}^2$	all $\alpha_i = 0$
$R_2(\alpha\|\mu,\beta,\gamma)$	same as $R_1(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i = 0$
$R_o(\alpha\|\mu,\beta)$	same as $R_1(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i = \bar{\alpha}_.$ | all $\gamma_{ij} = \bar{\gamma}_{..}$
$R_1(\alpha\|\mu,\beta)$	same as $R_1(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i = 0$ | all $\gamma_{ij} = 0$
$R_2(\alpha\|\mu,\beta)$	same as $R_1(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i = 0$ | all $\gamma_{ij} = 0$
$R_o(\alpha\|\mu)$	same as $R_1(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i + \bar{\gamma}_{i.}$ equal
$R_1(\alpha\|\mu)$	same as $R_1(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i = 0$
$R_2(\alpha\|\mu)$	same as $R_1(\alpha\|\mu,\beta,\gamma)$	all $\alpha_i = 0$